US 20110281429A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2011/0281429 A1

(43) Pub. Date: Nov. 17, 2011

Ganguly et al.

(54) **MULTI-LAYER CHARGE TRAP SILICON NITRIDE/OXYNITRIDE LAYER ENGINEERING WITH INTERFACE REGION CONTROL**

(75) Inventors: **Udayan Ganguly**, San Jose, CA (US); **Christopher S. Olsen**, Fremont, CA (US); **Sean M. Seutter**, San Jose, CA (US); **Lucien Date**, Ottignies (BE)

(73) Assignee: **Applied Materials, Inc.**, Santa Clara, CA (US)

(21) Appl. No.: **13/189,225**

(22) Filed: **Jul. 22, 2011**

Related U.S. Application Data

(62) Division of application No. 12/610,457, filed on Nov. 2, 2009.

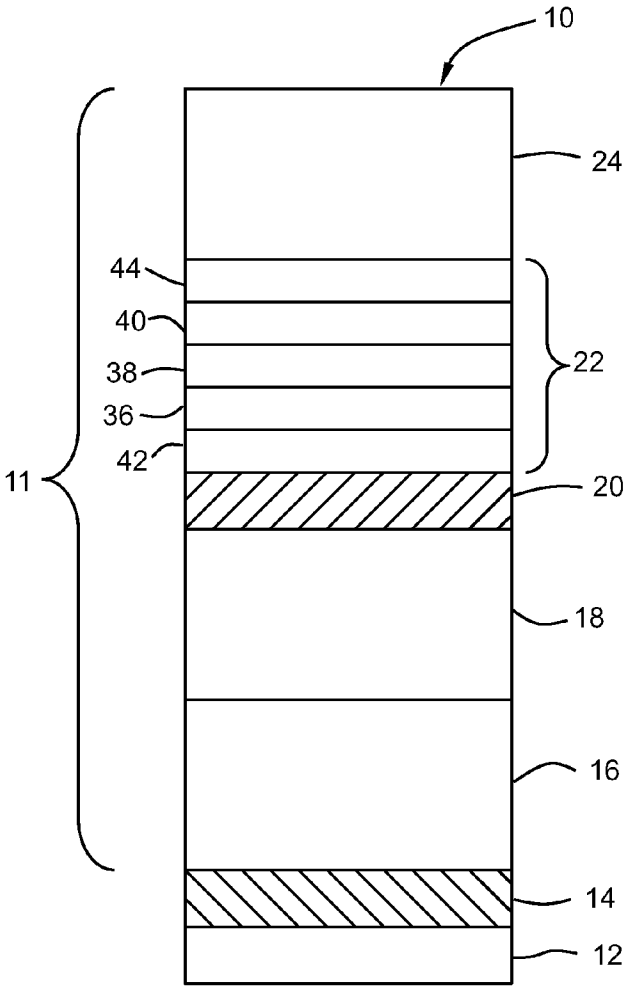
Publication Classification

(51) **Int. Cl.**
H01L 21/283 (2006.01)

(52) **U.S. Cl.** **438/591**; 257/E21.159

(57) **ABSTRACT**

A non-volatile memory semiconductor device comprising a semiconductor substrate having a channel and a gate stack above the channel. The gate stack comprises a tunnel layer adjacent to the channel, a charge trapping layer above the tunnel layer, a charge blocking layer above the charge trapping layer, a control gate above the charge blocking layer, and an intentionally incorporated interface region between the charge trapping layer and the charge blocking layer. The charge trapping layer comprises a compound including silicon and nitrogen, the charge blocking layer contains an oxide of a charge blocking component, and the interface region comprises a compound including silicon, nitrogen and the charge blocking component. The tunnel layer may comprise up to three tunnel sub-layers, the charge trapping layer may comprise two trapping sub-layers, and the charge blocking layer may comprise up to five blocking sub-layers. Various gate stack formation techniques can be employed.



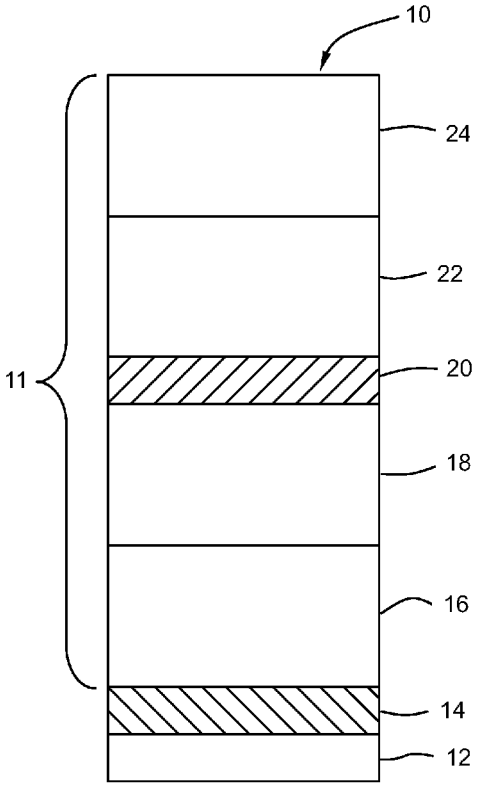


FIG. 1

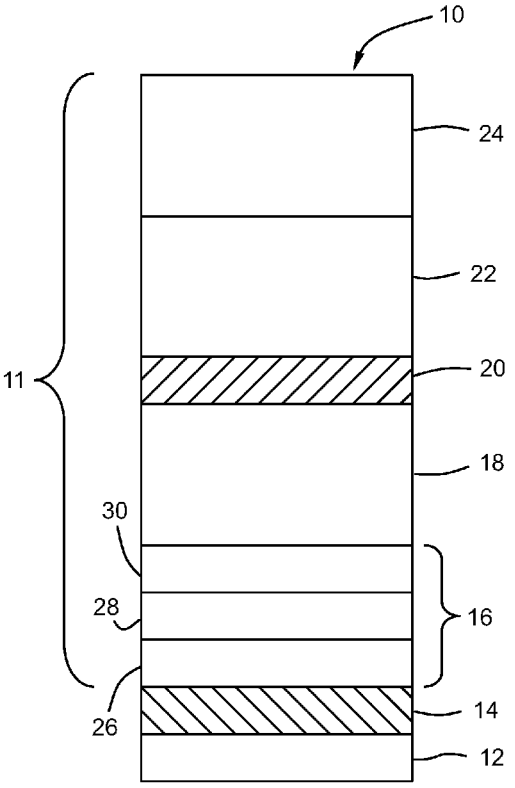


FIG. 2

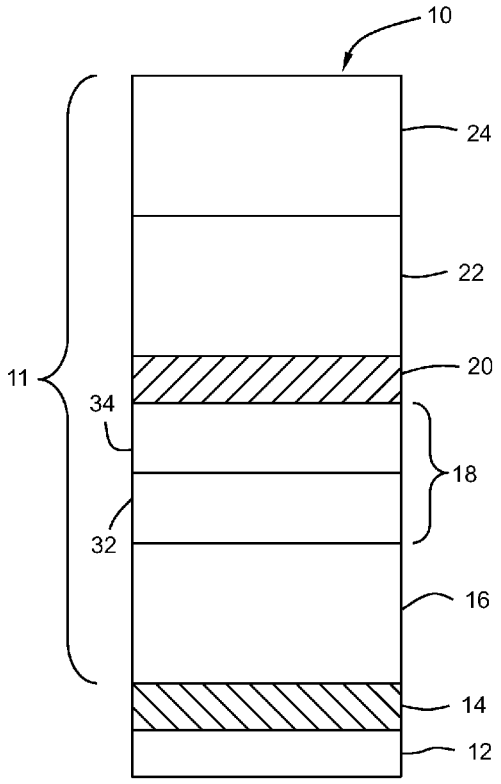


FIG. 3

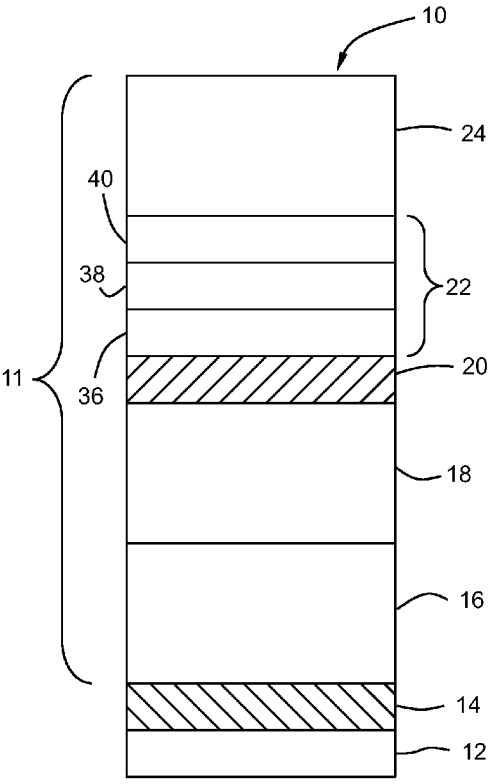


FIG. 4

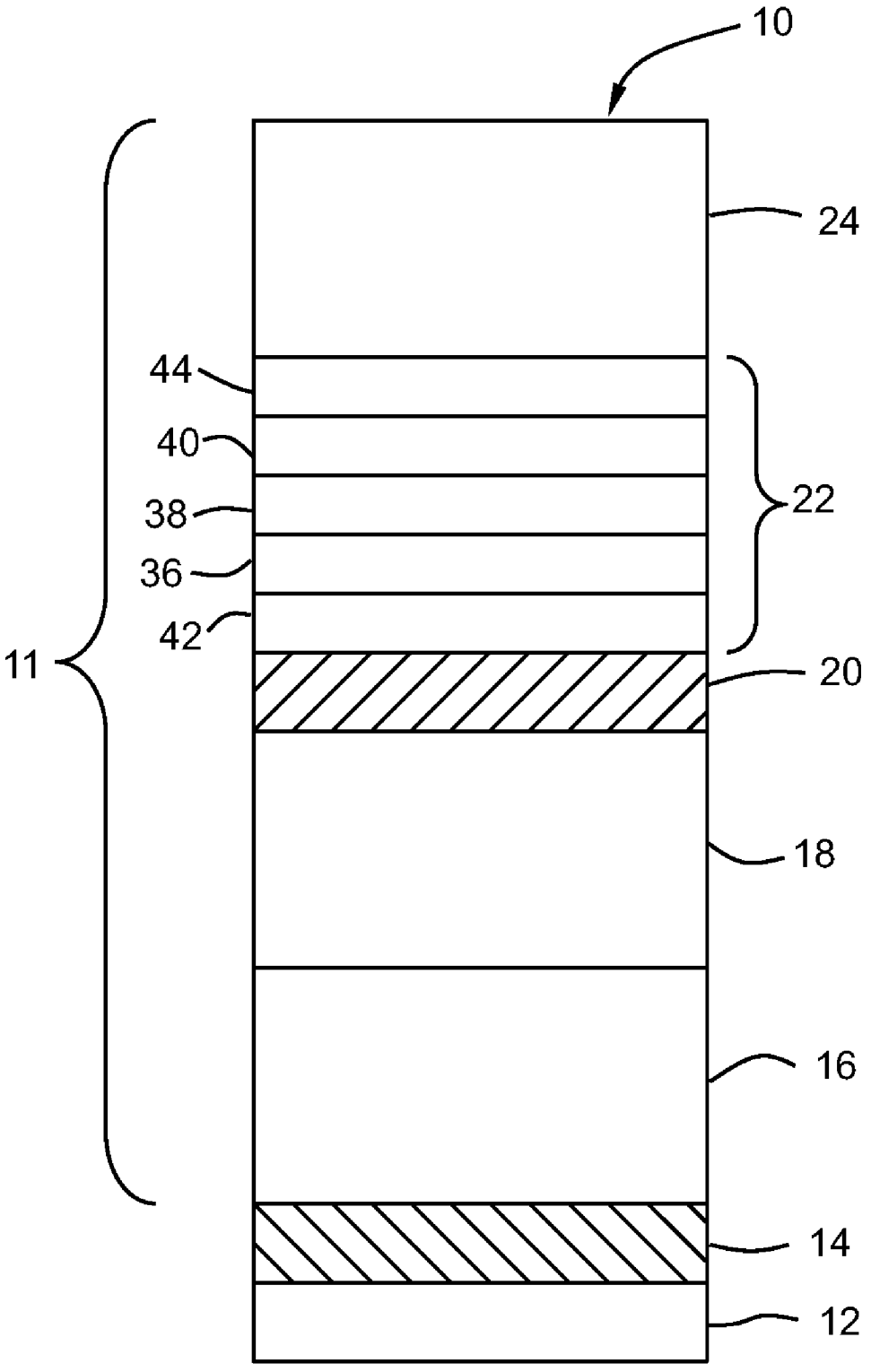


FIG. 5

MULTI-LAYER CHARGE TRAP SILICON NITRIDE/OXYNITRIDE LAYER ENGINEERING WITH INTERFACE REGION CONTROL

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a divisional of U.S. patent application Ser. No. 12/610,457, filed on Nov. 2, 2009, the contents of which is specifically incorporated by reference herein in its entirety.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] Embodiments of the present invention generally relate to non-volatile memory devices, and more particularly to a charge trap flash memory device that yields minimum endurance degradation and a maximum program/erase window.

[0004] 2. Description of the Related Art

[0005] Non-volatile memory (NVM) devices are electronic memory devices that retain their content when external power is removed. These devices are used in a wide variety of commercial and military electronic devices and equipment, such as hand-held telephones, radios and digital cameras. An NVM device comprises an insulating barrier, which can include multiple dielectric layers, located between a charge supply region and a charge storage region. The charge storage region can take the form of a floating gate structure or a charge trapping layer. Programming such NVM devices is accomplished by tunneling charge carriers of a first type such as electrons from the charge supply region, through the insulating barrier, toward the charge storage region.

[0006] One type of NVM device is flash memory. Flash memory devices can be further divided into floating gate flash memory and charge trap flash (CTF) memory. In floating gate memory devices, a floating gate of a metal layer or a metal-like layer is used as the charge storage layer. In CTF memory devices, such as in semiconductor-oxide-nitride-oxide-semiconductor (SONOS) memory devices, a charge trapping dielectric layer is used.

[0007] Essentially, a SONOS memory cell is a conventional n-channel metal-oxide-semiconductor transistor (NMOS) with a gate dielectric of a thermal oxide layer of about 2-7 nm thickness, a silicon nitride layer of about 3-12 nm thickness and a second oxide layer with a thickness of between 5 nm and 20 nm. The thin oxide layer is an exemplary tunneling layer and the nitride layer is an exemplary charge trapping layer. At a positive gate bias, electrons can tunnel from the substrate through the ultra thin oxide layer to the nitride layer, where they are subsequently trapped. Silicon nitride, for example, has an intrinsic property of trapping charge (electrons or holes).

[0008] In SONOS memory devices, the trapped charges create a threshold voltage shift of the transistor. The threshold voltage V_{th} varies depending on whether the NVM device is in the program state, wherein charges (electrons) are injected into the charge trapping layer, or in the erase state, wherein electrons leave (or holes enter) the charge trapping layer. This state in turn varies the gate voltage level needed to allow a certain current level to conduct through the channel. Hence, operation of charge trapping NVM semiconductor devices is based on the threshold voltage Vth being varied by charges trapped or stored in a charge trapping layer.

[0009] NVM devices that use charge trapping as a charge storage mechanism instead of a floating gate are becoming more and more prevalent. Such devices store charge in a charge trapping layer, such as a silicon nitride layer sandwiched between two oxide layers or, as an alternative, in nanocrystals. Charge trapping NVM devices are becoming more prevalent partially because they are believed to have considerable potential for use in future complementary metal-oxide-semiconductor generations, particularly for technologies with dimensions of 90 nm and smaller. One difference between charge trapping devices and floating gate devices is that with charge trapping devices electrons are trapped in energy minima caused by imperfections in the charge trapping layer or, in the case of nanocrystal memories, on nanocrystals embedded in a gate oxide. These energy minima act as localized charge storage sites that are isolated from each other in which charge is trapped and stored. In the case of electrons, for example, the free electron energy levels associated with such imperfections and/or nanocrystals are below the free electron energy levels of the surrounding material, thereby creating energy wells and trapping free electrons therein at the locations of the imperfections and/or nanocrystals.

[0010] As mentioned, one reason for the growing interest in charge trapping devices is that such devices are relatively easy to scale with associated reductions in physical geometries for future semiconductor processing technology generations. For instance, the use of charge trapping devices eliminates floating gate patterning issues, such as those related to lithography, overlay and topography. Moreover, charge trapping devices may be programmed and erased using lower voltages than floating gate devices implemented in the same semiconductor process. The ability to use lower voltages is important, especially in embedded memories, since the market continues to demand devices that use lower operating voltages and have reduced power consumption. Another requirement is endurance which needs to meet different specifications for different application. A stringent endurance requirement is needed for solid state drives which is a growth sector for NAND Flash, (see JEDEC specifications for NAND performance).

[0011] However, prior art charge trapping NVM devices have certain disadvantages. One such drawback relates to the relatively poor erasing efficiency characteristic of these memory devices that prevents them from meeting backwards compatibility requirements of a $\sim 10^{-4}$ second erasing time at a negative (around −3 V) threshold voltage V_{th}. The known charge trap flash memory devices have this problem because of electron back tunneling through the blocking layer. Such back tunneling causes the erase threshold voltage V_{th} to not drop off sufficiently or sufficiently quickly in known SONOS devices. For example, the erase threshold voltage V_{th} in known devices should typically drop from about 1 V to about −3 V during the desired erase time of about 10^{-3} seconds. The erase threshold voltage V_{th} may then continue to decrease or may increase, especially when the gate bias is about −17 V to −15 V.

[0012] Another disadvantage is the limited data retention capability of known charge trapping devices. Data retention is the ability of an NVM device to retain data programmed into individual memory cells. This limited data retention capability is due, in part, to the use of thin dielectrics between the

substrate (e.g., the charge supply region) and the charge trapping layer. While the use of a thicker conventional tunnel dielectric such as SiO_2 would improve this data retention capability, this improvement would come at the expense of worsened erase saturation for the devices and, consequently, the duration of a program/erase window (increased duration) for current charge trapping NVM devices. Erase saturation is the inability to completely remove or compensate for the charge stored in the charge storage region of a charge trapping NVM device after programming the device. Therefore, alternative approaches for implementing charge trapping NVM devices are desirable.

[0013] Retention and memory window (before and after cycling or endurance) are two challenges associated with CTF performance. Silicon nitride trap layer engineering by changing the composition (for example Si:N ratio) provides an important trade-off. As the atomic percent N in SiN_x:H (hydrogen is frequently incorporated from precursors in the range of about 0 to about 15 atomic percent) is increased two trends are observed, the memory window decreases and the retention is strongly enhanced. In addition it has been observed that the memory window degradation after endurance increases (worsens) with the increase in the nitrogen content.

[0014] The profitability of CTF technology is driven by multi-level cell (MLC) operation which requires a large memory window. On the other hand, retention is a core specification. Due to conflicting trends, a trade-off is present. Nitrogen-rich nitride can produce satisfactory retention but enhanced endurance memory window degradation. An ideal solution would provide satisfactory retention and improve endurance degradation for maximum memory window for reliable MLC operation.

[0015] Thus, there is a need for an optimal NVM semiconductor device that simultaneously provides minimum endurance degradation and a maximum program/erase window, and a corresponding manufacturing method.

SUMMARY OF THE INVENTION

[0016] One or more embodiments of the invention are directed to non-volatile memory semiconductor devices comprising a semiconductor substrate having a channel and a gate stack above the channel. The gate stack comprises a tunnel layer adjacent to the channel, a charge trapping layer above the tunnel layer comprising a compound including silicon and nitrogen, a charge blocking layer above the charge trapping layer containing an oxide of a charge blocking component, a control gate above the charge blocking layer, and an interface region between the charge trapping layer and the charge blocking layer. The interface region is intentionally incorporated into the gate stack and comprises a compound including silicon, nitrogen and the charge blocking component.

[0017] In detailed embodiments, the semiconductor substrate comprises silicon. In more detailed embodiments, the tunnel layer comprises at least one compound selected from SiO_2, Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HIO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$ and $LaAlO_3$. In further detailed embodiments, the tunnel layer comprises at least one silicon nitride compound having a chemical formula of Si_xN_y. Hydrogen is another element found in deposited SiN which affects traps and trap generation. Hydrogen is incorporated into the silicon nitride from the chemical precursors used and is not unexpected. The hydrogen is frequently present in the range of about 0 to about 15 atomic percent. $Si_xN_yH_z$ is another possibility where H may be 0-20 atomic percentage of the composition. In specific embodiments, the tunnel layer comprises SiON.

[0018] In some detailed embodiments, the tunnel layer comprises a first tunnel sub-layer, a second tunnel sub-layer and a third tunnel sub-layer. The first tunnel sub-layer comprises at least one compound selected from the group consisting of SiO_2 and SiON.

[0019] The second tunnel sub-layer comprises at least one compound selected from the group consisting of Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HfO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$, $LaA1O_3$ and Si_xN_y. The third tunnel sub-layer comprises at least one compound selected from the group consisting of SiO_2 and SiON. According to specific embodiments, the tunnel layer has a thickness in the range of about 5 nm to 12 nm.

[0020] The charge trapping layer of some embodiments comprises a first trapping sub-layer and a second trapping sub-layer. The first trapping sub-layer is adjacent to the tunnel layer and has a first composition engineered to provide a low charge trap density to increase retention of information. The second trapping sub-layer has a second composition adjacent to the interface region.

[0021] In detailed embodiments, the first trapping sub-layer comprises stoichiometric silicon nitride. In specific embodiments, the first trapping sub-layer comprises a nitrogen-rich silicon nitride compound containing about the same as or more nitrogen on an atomic percentage basis than is present in stoichiometric silicon nitride. In some detailed embodiments, the first trapping sub-layer comprises a composition of at least one silicon nitride compound having a chemical formula of Si_xN_y present in the range of about 50% to 100% on a atomic percentage basis, and oxygen present in the range of about 0% to 50% on a atomic percentage basis.

[0022] In detailed embodiments, the second trapping sub-layer is engineered to provide a low charge trap generation and minimize endurance degradation of the device. In some embodiments, the second trapping sub-layer has a high charge trap density for improved memory window. In specific embodiments, the second trapping sub-layer comprises a silicon-rich silicon nitride compound containing more silicon on a atomic percentage basis than is present in stoichiometric silicon nitride.

[0023] In one or more embodiments, the charge trapping layer is a graded layer having a variable composition and a thickness in the range of about 1 nm to 20 nm The charge trapping layer comprises nitrogen-rich Si_xN_y adjacent to the tunnel layer and silicon-rich Si_xN_y adjacent to the interface region.

[0024] In some embodiments, the interface region comprises SiON and has a thickness in the range of about 1 nm to 5 nm.

[0025] In detailed embodiments, the charge blocking layer comprises a first blocking sub-layer, a second blocking sub-layer and a third blocking sub-layer. The first blocking sub-layer comprises at least one compound selected from the group consisting of SiO_2 and SiON. The second blocking sub-layer comprises at least one compound selected from the group consisting of Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$, $LaA1O_3$ and Si_xN_y. The third blocking sub-layer comprises at least one compound selected from the group consisting of SiO_2 and SiON.

[0026] In specific embodiments, the charge blocking layer further comprises at least a fourth blocking sub-layer and a fifth blocking sub-layer. The fourth blocking sub-layer adjacent to the interface region, the fourth blocking sub-layer comprises at least one silicon nitride compound having a chemical formula of Si_xN_y. The fifth blocking sub-layer is adjacent to the control gate and comprises at least one silicon nitride compound having a chemical formula of Si_xN_y.

[0027] In detailed embodiments, the charge blocking layer has a thickness in the range of about 10 nm to 15 nm.

[0028] Additional embodiments of the invention are directed to methods of forming a gate stack on a semiconductor substrate having a channel. A tunnel layer is deposited over the channel. A charge trapping layer comprising a compound including silicon and nitrogen is deposited on top of the tunnel layer. A charge blocking layer containing an oxide of a charge blocking component is deposited on top of an interface region. An interface region is formed between the charge trapping layer and the charge blocking layer. The interface region comprises a compound including silicon, nitrogen and the charge blocking component. A control gate is placed on top of the charge blocking layer.

[0029] In detailed embodiments, depositing a tunnel layer over the channel comprises depositing a first tunnel sub-layer, a second tunnel sub-layer and a third tunnel sub-layer. The first tunnel sub-layer is deposited over the channel. The first tunnel sub-layer comprises at least one component selected from the group consisting of SiO_2 and SiON.

[0030] The second tunnel sub-layer is deposited on top of the first tunnel sub-layer. The second tunnel sub-layer comprises at least one component selected from the group consisting of Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HfO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$, $LaAlO_3$ and Si_xN_y. A third tunnel sub-layer is deposited on top of the second tunnel sub-layer. The third tunnel sub-layer comprises at least one component selected from the group consisting of SiO_2 and SiON.

[0031] In some detailed embodiments, depositing a charge trapping layer comprises depositing a first trapping sub-layer and a second trapping sub-layer. The first trapping sub-layer is deposited in a first chamber without air exposure. The first trapping sub-layer being adjacent to the tunnel layer. The second trapping sub-layer is deposited in the first chamber without air exposure. The second trapping sub-layer being adjacent to the interface region.

[0032] In specific embodiments, forming the interface region comprises one or more of controlled oxidation of the second trapping sub-layer, oxidizing the second trapping sub-layer by exposing the second trapping sub-layer to air, and depositing a discrete interface region on top of the charge trapping layer in the first chamber without air exposure.

[0033] In further specific embodiments, controlled oxidation of the second trapping sub-layer comprises one or more of oxidizing the second trapping sub-layer using wet oxidation, oxidizing the second trapping sub-layer using dry oxidation, oxidizing the second trapping sub-layer using radical oxidation, oxidizing the second trapping sub-layer using plasma oxidation, and oxidizing the second trapping sub-layer using oxygen implantation.

[0034] In other specific embodiments, the first trapping sub-layer, the second trapping sub-layer and the interface region are deposited in separate chambers without air exposure between depositions.

[0035] The foregoing has outlined rather broadly certain features and technical advantages of the present invention. It should be appreciated by those skilled in the art that the specific embodiments disclosed may be readily utilized as a basis for modifying or designing other structures or processes within the scope present invention. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope of the invention as set forth in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] So that the manner in which the above recited features of the present invention can be understood in detail, a more particular description of the invention, briefly summarized above, may be had by reference to embodiments, some of which are illustrated in the appended drawings. It is to be noted, however, that the appended drawings illustrate only typical embodiments of this invention and are therefore not to be considered limiting of its scope, for the invention may admit to other equally effective embodiments.

[0037] FIG. **1** illustrates an enlarged front elevation view of one embodiment of the non-volatile memory semiconductor device according to the present invention.

[0038] FIG. **2** illustrates an enlarged front elevation view of another embodiment of the non-volatile memory semiconductor device according to the present invention.

[0039] FIG. **3** illustrates an enlarged front elevation view of an additional embodiment of the non-volatile memory semiconductor device according to the present invention.

[0040] FIG. **4** illustrates an enlarged front elevation view of an additional embodiment of the non-volatile memory semiconductor device according to the present invention.

[0041] FIG. **5** illustrates an enlarged front elevation view of an additional embodiment of the non-volatile memory semiconductor device according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0042] An embodiment of the non-volatile memory semiconductor device **10** is depicted in FIG. **1**. The device **10** includes a semiconductor substrate **12** having a channel **14** and a gate stack **11** above the channel **14**. As is known in the art, the substrate **12** may comprise silicon. The gate stack **11** comprises a tunnel layer **16** adjacent to the channel **14**, a charge trapping layer **18** above the tunnel layer **16**, a charge blocking layer **22** above the charge trapping layer **18**, a control gate **24** above the charge blocking layer **22**, and an interface region **20** between the charge trapping layer **18** and the charge blocking layer **22**.

[0043] The charge trapping layer **18** comprises a compound containing silicon and nitrogen, while the charge blocking layer **22** contains an oxide of a charge blocking component. The interface region **20** comprises a compound containing silicon, nitrogen and a charge blocking component. Since the interface region **20** is intentionally incorporated into the gate stack **11**, it can be engineered for a desired rate of formation, thickness and composition, for example. Thus, the interface region **20** is generally distinct from, and preferable to, any products resulting from a spontaneous oxidation reaction of the charge trapping layer **18** and atmospheric air.

[0044] While dimensions of the tunnel layer **16** can vary, the thickness is generally in the range of about 5 nm to 12 nm. Examples of suitable compositions of the tunnel layer **16** include, but are not limited to, silicon oxides, silicon nitrides,

silicon oxynitrides, materials having large dielectric constants, and combinations thereof. In detailed embodiments the tunnel layer **16** comprises one or more compounds having chemical formulae of SiO_2, Si_xN_y, SiON, Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HfO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$ and $LaA1O_3$ may be utilized in the tunnel layer **16**.

[0045] In certain embodiments of the device **10**, as illustrated in FIG. **2**, the tunnel layer **16** may comprise a first tunnel sub-layer **26**, a second tunnel sub-layer **28** and a third tunnel sub-layer **30**. In specific embodiments, the first tunnel sub-layer **26** and the third tunnel sub-layer **30** each comprise SiO_2, SiON or both. In detailed embodiments, the second tunnel sub-layer **28** comprises at least one compound selected from the group consisting of Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HfO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$, $LaAlO_3$ and Si_xN_y. Other materials that have large dielectric constants may also be incorporated into the second tunnel sub-layer **28**.

[0046] The second tunnel sub-layer **26** may be optimized for retention, implying smaller trap density and deeper traps. In various embodiments, the second tunnel sub-layer **26** is a stoichiometric SiN deposited by chemical vapor deposition (CVD), a nitrogen-rich SiN deposited by CVD, a SiN with 0-50% oxygen uniformly deposited by CVD or a SiN with 0-50% oxygen implanted by DPO or P3i. The thickness of the second tunnel sub-layer **26** in some embodiments is in the range of about 1 to about 10 nm

[0047] In specific embodiments, the combination of the charge trapping layer **18** and the interface region **20** may be optimized for low trap generation during endurance and the SiN layer may be optimized with high trap density for memory window improvement. Controlling the composition of the layers and thicknesses may impact the performance of resulting devices.

[0048] In order to optimize the device **10**, the charge trapping layer **18** may be divided into regions that each provide increased retention of information, decreased endurance degradation or expanded memory window. FIG. **3** shows such an embodiment. The charge trapping layer **18** comprises a first trapping sub-layer **32** adjacent to the tunnel layer **16** and a second trapping sub-layer **34** adjacent to the interface region **20**. The first trapping sub-layer **32** has a first composition engineered to provide a low charge trap density for increasing retention of information. On the other hand, the second trapping sub-layer **34** can be designed in one of two ways: with a low charge trap generation for decreased endurance degradation or with a high charge trap density for improved memory window. In certain aspects of the invention, the first trapping sub-layer **32** comprises stoichiometric silicon nitride or a nitrogen-rich silicon nitride compound containing more nitrogen on an atomic percentage basis than is present in stoichiometric silicon nitride.

[0049] According to detailed embodiments of the device **10**, the first trapping sub-layer **32** comprises a composition of at least one silicon nitride compound having a chemical formula of Si_xN_y present in the range of about 50% to 100% on an atomic percentage basis and oxygen present in the range of about 0% to 50% on the same basis. The second trapping sub-layer **34** may comprise a silicon-rich silicon nitride compound containing more silicon on an atomic percentage basis than is present in stoichiometric silicon nitride.

[0050] Alternatively, as represented in FIG. **2**, the charge trapping layer **18** may be a single graded layer that has a variable composition. The charge trapping layer **18** generally comprises nitrogen-rich Si_xN_y adjacent to the tunnel layer **16** and silicon-rich Si_xN_y adjacent to the interface region **20**. The thickness of the charge trapping layer **18** may be in the range of about 1 nm to about 20 nm. For the following reasons, this embodiment is believed to be more useful than depicted in FIG. **3**. When the composition of the charge trapping layer is changed, a sharp offset in conduction band and valence band results. This discontinuity in conduction band or valence band incites high-energy electron scattering at the interface between the first trapping sub-layer **32** and the second trapping sub-layer **34** during the repeated program/erase operation or program/erase cycling or endurance. The electronic energy is transferred to the dielectric network at the interface, which may generate defects such as broken bonds during operation. Undesirable modifications of the device **10** and its performance throughout its lifetime of program/erase cycling or endurance result. By utilizing the graded charge trapping layer **18**, the abruptness of the interface is reduced, the scattering is reduced or even avoided, and performance of the device **10** is improved. With this configuration, scattering is postponed and occurs when electrons reach the interface between the interface region **20** and the charge blocking layer **22**, which is specially engineered for low defect generation under such conditions. Preferably, the interface region **20** comprises SiON and has a thickness in the range of about 1 nm to 5 nm

[0051] As discussed previously with respect to FIG. **1**, the charge blocking layer **22** may comprise a single layer. In other embodiments of the invention, as shown in FIG. **4**, the charge blocking layer **22** may comprise a first blocking sub-layer **36**, a second blocking sub-layer **38** and a third blocking sub-layer **40**. In detailed embodiments, the first blocking sub-layer **36** and the third blocking sub-layer **40** may each comprise SiO_2, SiON or both. The second blocking sub-layer **38** may comprise at least one compound selected from the group consisting of Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HfO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$, $LaAlO_3$ and Si_xN_y. Other materials that have large dielectric constants may also be incorporated into the second blocking sub-layer **38**.

[0052] Alternatively, as illustrated in FIG. **5**, the charge blocking layer **22** may comprise a first blocking sub-layer **36**, a second blocking sub-layer **38**, a third blocking sub-layer **40**, a fourth blocking sub-layer **42** and a fifth blocking sub-layer **44**. The fourth blocking sub-layer **42** is adjacent to the interface region **20** while the fifth blocking sub-layer **44** is adjacent to the control gate **24**. Preferably, the fourth blocking sub-layer **42** and the fifth blocking sub-layer **44** each comprises at least one silicon nitride compound having a chemical formula of Si_xN_y. The charge blocking layer **22** has a total thickness in the range of about 10 nm to 15 nm. In specific embodiments incorporating multilayered blocking oxide layers, each layer can have a thickness in the range of about 1 nm to about 5 nm.

[0053] Methods used to form the gate stack on the substrate **12** will now be described. The tunnel layer **16** is deposited on the substrate **12** above the channel **14**. The charge trapping layer **18** is deposited on top of the tunnel layer **16**. In addition, the interface region **20** is formed on top of the charge trapping layer **18**. The charge blocking layer **22** is deposited on top of the interface region **20**. A control gate is placed on top of the charge blocking layer **22**. For embodiments in which the tunnel layer comprises multiple tunnel sub-layers, the first tunnel sub-layer **26** is deposited over the channel **14** first.

Then, the second tunnel sub-layer **28** is deposited on top of the first tunnel sub-layer **26**. The third tunnel sub-layer **30** is deposited on top of the second tunnel sub-layer **28**. For embodiments in which the charge trapping layer **18** comprises multiple trapping sub-layers, the first trapping sub-layer **32** is deposited on the tunnel layer **16** in a first chamber without air exposure. Subsequently, the second trapping sub-layer **34** is deposited on top of the first trapping sub-layer **32**.

[0054] Importantly, formation of the interface region **20** can be accomplished in several ways. The second trapping sub-layer **34** can be oxidized in a controlled chemical reaction. Various techniques can be employed in this regard, including wet oxidation, dry oxidation, radical oxidation, plasma oxidation and oxygen implantation. Other suitable controlled processes known in the art can be used as well. Alternatively, the second trapping sub-layer **34** can be oxidized by exposing it to air. The interface region **20** can also be deposited on top of the charge trapping layer **18** in the first chamber without air exposure. Preferably, the second trapping sub-layer **34** and the interface region **20** are deposited in separate chambers without air exposure between depositions. The oxidation may function as a method to add oxygen or additionally to partially remove or modulate nitrogen and hydrogen. In detailed embodiments, the interface region **20** is formed by one or more of controlled oxidation of the second trapping sub-layer **34**, oxidizing the second trapping sub-layer **34** by exposing the second trapping sub-layer **34** to air and depositing a discrete interface region **20** on top of the charge trapping layer **18** in the first chamber without air exposure.

[0055] Reference throughout this specification to "one embodiment," "certain embodiments," "one or more embodiments" or "an embodiment" means that a particular feature, structure, material, or characteristic described in connection with the embodiment is included in at least one embodiment of the invention. Thus, the appearances of the phrases such as "in one or more embodiments," "in certain embodiments," "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily referring to the same embodiment of the invention. Furthermore, the particular features, structures, materials, or characteristics may be combined in any suitable manner in one or more embodiments. The order of description of the above method should not be considered limiting, and methods may use the described operations out of order or with omissions or additions.

[0056] It is to be understood that the above description is intended to be illustrative, and not restrictive. Many other embodiments will be apparent to those of ordinary skill in the art upon reviewing the above description. The scope of the invention should, therefore, be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

1. A method of forming a gate stack on a semiconductor substrate having a channel comprising:

depositing a tunnel layer over the channel;

depositing a charge trapping layer comprising a compound including silicon and nitrogen on top of the tunnel layer, the charge trapping layer comprising at least a first trapping sub-layer adjacent to the tunnel layer, the first trapping sub-layer having a first composition engineered to provide a low charge trap density to increase retention of information, the first trapping sub-layer comprising a composition of at least one silicon nitride compound having a chemical formula of Si_xN_y present in the range of about 50% to 100% on an atomic percentage basis and oxygen present in the range of about 0% to about 50% on an atomic percentage basis;

depositing a charge blocking layer containing an oxide of a charge blocking component on top of an interface region;

forming the interface region between the charge trapping layer and the charge blocking layer, the interface region comprising a compound including silicon, nitrogen and the charge blocking component; and

placing a control gate on top of the charge blocking layer.

2. The method of forming a gate stack of claim **1**, wherein depositing a tunnel layer over the channel comprises:

depositing a first tunnel sub-layer over the channel, the first tunnel sub-layer comprising at least one component selected from the group consisting of SiO_2 and SiON;

depositing a second tunnel sub-layer on top of the first tunnel sub-layer, the second tunnel sub-layer comprising at least one component selected from the group consisting of Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HfO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$, $LaAlO_3$ and Si_xN_y; and

depositing a third tunnel sub-layer on top of the second tunnel sub-layer, the third tunnel sub-layer comprising at least one component selected from the group consisting of SiO_2 and SiON.

3. The method of forming a gate stack of claim **1**, wherein depositing a charge trapping layer comprises:

depositing a first trapping sub-layer in a first chamber without air exposure, the first trapping sub-layer adjacent to the tunnel layer; and

depositing a second trapping sub-layer in the first chamber without air exposure, the second trapping sub-layer adjacent to the interface region.

4. The method of forming a gate stack of claim **3**, wherein forming the interface region comprises one or more of controlled oxidation of the second trapping sub-layer, oxidizing the second trapping sub-layer by exposing the second trapping sub-layer to air, and depositing a discrete interface region on top of the charge trapping layer in the first chamber without air exposure.

5. The method of claim **4**, wherein controlled oxidation of the second trapping sub-layer comprises one or more of oxidizing the second trapping sub-layer using wet oxidation, oxidizing the second trapping sub-layer using dry oxidation, oxidizing the second trapping sub-layer using radical oxidation, oxidizing the second trapping sub-layer using plasma oxidation, and oxidizing the second trapping sub-layer using oxygen implantation.

6. The method of forming a gate stack of claim **4**, wherein the first trapping sub-layer, the second trapping sub-layer and the interface region are deposited in separate chambers without air exposure between depositions.

7. The method of forming a gate stack of claim **1**, wherein the step of depositing a tunnel layer over the channel comprises a tunnel layer having at least one compound selected from SiO_2, Al_2O_3, MgO, SrO, BaO, TiO, Ta_2O_5, $BaTiO_3$, $BaZrO_3$, ZrO_2, HfO_2, Y_2O_3, $ZrSiO_4$, $HfSiO_4$ and $LaAlO_3$.

8. The method of forming a gate stack of claim **1**, wherein the step of depositing a tunnel layer over the channel comprises a tunnel layer having at least one silicon nitride compound having a chemical formula of $Si_xN_yH_z$.

9. The method of forming a gate stack of claim **1**, wherein the step of depositing a tunnel layer over the channel comprises a tunnel layer having SiON.

10. The method of forming a gate stack of claim **1**, wherein the step of depositing a tunnel layer over the channel comprises a tunnel layer having a thickness in the range of about 5 nm to 12 nm.

11. The method of forming a gate stack of claim **1**, wherein the step of forming the interface region between the charge trapping layer and the charge blocking layer comprises an interface region having a thickness in the range of about 1 nm to 5 nm.

12. The method of forming a gate stack of claim **1**, wherein the step of depositing a charge blocking layer comprises a charge blocking layer having a thickness in the range of about 10 nm to 15 nm.

* * * * *